CLUSTERING DOKUMEN TEKS BERDASARKAN *FINGERPRINT BIWORD WINNOWING* DENGAN MENGGUNAKAN METODE *K-MEANS*

TUGAS AKHIR

Diajukan Sebagai Salah Satu
Syarat Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika

Oleh

YULISKA
10951007997

FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2014

LEMBAR PENGESAHAN

CLUSTERING DOKUMEN TEKS BERDASARKAN *FINGERPRINT BIWORD WINNOWING* DENGAN MENGGUNAKAN METODE *K-MEANS*

TUGAS AKHIR

Oleh:

YULISKA
10951007997

Telah dipertahankan di depan sidang dewan penguji
Sebagai salah satu syarat untuk memperoleh gelar sarjana Teknik Informatika
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
di Pekanbaru, pada tanggal 26 Juni 2014

Pekanbaru, 26 Juni 2014
Mengesahkan

Dekan	Ketua Jurusan
Hj. Helvi Yenita Morena, M.Si	Elin Haerani, S.T, M.Kom
NIP 19601125 198503 2 002	NIP. 19810523 200710 2 003

DEWAN PENGUJI :

Ketua	**: Surya Agustian, S.T, M.Kom**
Sekretaris	**: Surya Agustian, S.T, M.Kom**
Anggota I	**: Jasril, S.Si, M.Sc**
Anggota II	**: Elvia Budianita, S.T, M.Cs**

CLUSTERING DOKUMEN TEKS BERDASARKAN *FINGERPRINT BIWORD WINNOWING* DENGAN MENGGUNAKAN METODE *K-MEANS*

YULISKA
10951007997

Tanggal Sidang : 26 Juni 2014

Periode Wisuda :

Jurusan Teknik Informatika

Fakultas Sains Dan Teknologi

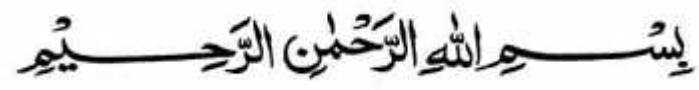
Univesitas Islam Negeri Sultan Syarif Kasim Riau

ABSTRAK

Kebutuhan akan informasi yang semakin meningkat menyebabkan informasi yang berbentuk dokumen teks mengalami penumpukan, sehingga proses pencarian menjadi sulit. Untuk itu, dilakukan suatu teknik *clustering* agar dokumen-dokumen yang memiliki tingkat kesamaan tinggi dapat berada pada satu kluster/kelompok yang sama, sehingga proses pencarian menjadi lebih mudah. Salah satu metode yang sering digunakan dalam kasus *clustering* adalah *K-Means*. Penelitian ini mencoba melakukan *clustering* terhadap dokumen teks dengan menggunakan metode *K-Means* berdasarkan *fingerprint biword winnowing* sebagai *feature* atau ciri dokumen teks, dimana *biword winnowing* adalah algoritma yang biasa dipakai untuk mendeteksi kesamaan isi suatu dokumen teks. Proses *clustering* diawali dengan mengekstrak *fingerprint* dari masing-masing dokumen yang ada pada koleksi dokumen dengan menggunakan algoritma *biword winnowing*. Kemudian dilakukan proses pembentukan dimensi dokumen yang merepresentasikan frekuensi kemunculan *fingerprint* dari masing-masing dokumen tersebut, dan dilanjutkan dengan proses *clustering* dokumen menggunakan metode *K-Means*. Selanjutnya, untuk mengetahui kualitas hasil pengelompokan dilakukan pengujian dengan menggunakan parameter *precision*. Dari keseluruhan hasil pengujian, kualitas hasil pengelompokan terhadap koleksi dokumen uji oleh aplikasi *clustering* dokumen teks ini dinilai baik dengan nilai *precision* 60%-88.33%.

Kata Kunci : *Biword Winnowing*, *Clustering*, *Fingerprint*, *K-Means*, *Precision*.

KATA PENGANTAR



Segala puji bagi Allah, Pencipta mausia dan alam semesta karena senantiasa melimpahkan rahmat dan karunianya yang tidak akan sanggup untuk penulis hitung, sehingga penulis dapat menyelesaikan laporan tugas akhir perkuliahan ini dengan baik. Shalawat dan salam rindu yang terdalam semoga senantiasa tercurah dan tersampaikan kepada Nabi Muhammad, keluarganya, para sahabatnya dan para pengikutnya yang setia hingga akhir zaman.

Tugas akhir dengan judul "*Clustering* dokumen teks berdasarkan *fingerprint biword winnowing* dengan menggunakan metode *K-Means*" ini penulis susun dengan maksud untuk memenuhi tugas akhir perkuliahan yang dibebankan kepada penulis di samping untuk menambah dan memperdalam kemampuan penulis dalam memahami materi perkuliahan.

Dalam kesempatan ini penulis juga ingin mengucapkan terima kasih kepada seluruh pihak yang telah membantu penulis dalam menyelesaikan laporan tugas akhir ini. Untuk itu penulis mengucapkan terima kasih yang tak terhingga kepada :

1. Allah SWT yang selalu melimpahkan rahmat dan nikmat-Nya kepada penulis hingga terselesaikannya Tugas Akhir ini.
2. Bapak Prof. DR. H. Munzir Hitami, MA, selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
3. Ibu Dra. Hj.Yenita Morena, M.Si, selaku Dekan Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau
4. Ibu Maylisda dan Ayah Zakaria tercinta, yang telah mendo'akan dan memberikan dukungan yang sangat luar biasa kepada penulis dalam menyelesaikan tugas akhir ini. Semoga beliau selalu dalam lindungan Allah SWT serta segala ketulusan dan pengorbanan beliau diridhoi oleh Allah SWT. Amin.
5. Ibu Elin Haerani, S.T, M.Kom, selaku Ketua Jurusan Teknik Informatika.

6. Bapak Novriyanto, S.T, M.Sc, selaku Pembimbing Akademis Penulis. Terima kasih untuk segala bimbingan dan masukan yang telah diberikan kepada penulis.
7. Bapak Surya Agustian, S.T, M.Kom, selaku Pembimbing tugas akhir Penulis yang selalu sabar dan meluangkan waktu untuk membimbing, memberikan saran dan kritik dalam penyusunan tugas akhir ini.
8. Bapak Jasril, S.Si, M.Sc, selaku Penguji I tugas akhir. Terima kasih untuk segala masukan yang telah diberikan kepada penulis demi kesempurnaan laporan tugas akhir ini.
9. Ibu Elvia Budianita, S.T, M.Cs, Selaku Penguji II tugas akhir. Terima kasih untuk segala masukan yang telah diberikan kepada penulis demi kesempurnaan laporan tugas akhir ini.
10. Bapak Muhammad Affandes, S.T, M.T, selaku Koordinator Tugas akhir Jurusan Teknik Informatika.
11. Adik-adikku Rully Dwi Andika dan Febila Anzari yang telah memberikan dorongan dan semangat dalam menyelesaikan tugas akhir ini.
12. Seluruh keluarga besar Penulis, yang selalu memberikan semangat untuk menyelesaikan tugas akhir ini.
13. Seluruh dosen dan staf Fakultas Sains dan Teknologi khususnya pada Jurusan Teknik Informatika. Terima kasih atas ilmu yang telah diberikan.
14. Teman-teman seperjuangan penulis di UIN Suska Riau, teman-teman TIF 2009, Khususnya TIF C. Semoga yang telah lulus mendapat apa yang dicita-citakan dan yang belum lulus diberi kemudahan oleh Allah untuk menyusul teman-teman yang telah lulus. Amin.
15. Seluruh sahabat penulis, terima kasih untuk segala dukungannya.
16. Seluruh pihak yang belum penulis cantumkan, terima kasih atas dukungannya kepada penulis.

Selanjutnya, penulis menyadari bahwa dalam penulisan laporan tugas akhir ini masih terdapat kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi terciptanya laporan yang lebih baik di masa yang akan datang melalui email penulis yuliska_mz@yahoo.com atau hatake.bee@gmail.com. Penulis berharap laporan ini dapat bermanfaat bagi para

pembaca, baik dari pihak UIN SUSKA RIAU maupun dari kalangan akademis serta masyarakat umum.

Pekanbaru, Juni 2014

Penulis

DAFTAR ISI

Halaman